# DTM and LDA

Joost Bloos

04/11/2021

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

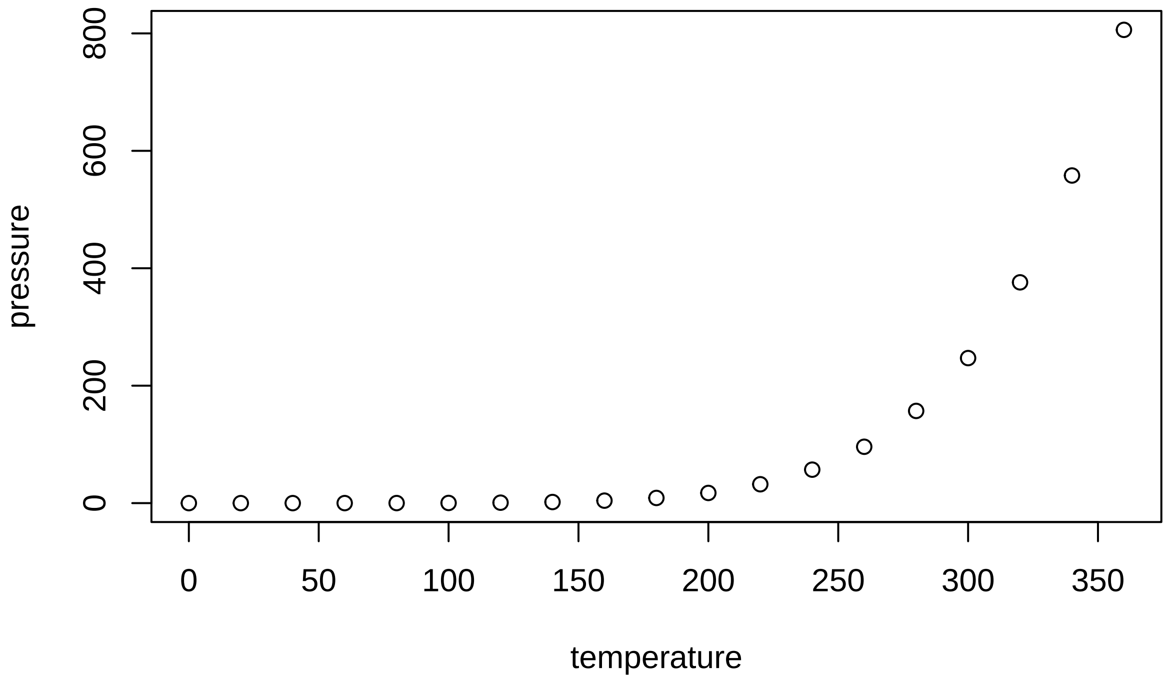
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00  
```

## Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

#{r setup, include=FALSE} #knitr::opts_chunk$set(echo = TRUE)

```
#install.packages("lexicon", dependencies = TRUE)
#install.packages("tm", dependencies = TRUE)
#install.packages("RWeka", dependencies = TRUE)
#install.packages("textstem", dependencies = TRUE)
#install.packages("textclean", dependencies = TRUE)
#install.packages("dplyr")
#install.packages("quanteda")
#install.packages("textstem")
#install.packages("text2vec")
#install.packages("namespace")
#install.packages("stopwords")
#install.packages("pairheatmap")
#install.packages("LDAvis")
#install.packages("servr")


#Loading the packages to the current workspace
lstPackages <- c('lexicon','tm','RWeka','textstem','textclean')
lapply(lstPackages, library, character.only = TRUE)
```

```
## Loading required package: NLP
## Loading required package: koRpus.lang.en

## Loading required package: koRpus

## Loading required package: sylly

## For information on available language packages for 'koRpus', run
##
##   available.koRpus.lang()
##
## and see ?install.koRpus.lang()

##
## Attaching package: 'koRpus'

## The following object is masked from 'package:tm':
##
##     readTagged

## [[1]]
## [1] "lexicon"   "stats"     "graphics"  "grDevices" "utils"     "datasets"
## [7] "methods"   "base"
##
## [[2]]
##  [1] "tm"        "NLP"       "lexicon"   "stats"     "graphics"  "grDevices"
##  [7] "utils"     "datasets"  "methods"   "base"
##
## [[3]]
##  [1] "RWeka"     "tm"        "NLP"       "lexicon"   "stats"     "graphics"
##  [7] "grDevices" "utils"     "datasets"  "methods"   "base"
##
## [[4]]
##  [1] "textstem"       "koRpus.lang.en" "koRpus"         "sylly"
##  [5] "RWeka"          "tm"             "NLP"            "lexicon"
##  [9] "stats"          "graphics"       "grDevices"      "utils"
## [13] "datasets"       "methods"        "base"
##
## [[5]]
##  [1] "textclean"      "textstem"       "koRpus.lang.en" "koRpus"
##  [5] "sylly"          "RWeka"          "tm"             "NLP"
##  [9] "lexicon"        "stats"          "graphics"       "grDevices"
## [13] "utils"          "datasets"       "methods"        "base"
```

```
library(quanteda)
```

```
## Package version: 3.1.0
## Unicode version: 13.0
## ICU version: 69.1

## Parallel computing: 4 of 4 threads used.

## See https://quanteda.io for tutorials and examples.
```

```
##
## Attaching package: 'quanteda'

## The following objects are masked from 'package:koRpus':
##
##     tokens, types

## The following object is masked from 'package:tm':
##
##     stopwords

## The following objects are masked from 'package:NLP':
##
##     meta, meta<-
```

```
library(stringr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(wordcloud)
```

```
## Loading required package: RColorBrewer
```

```
library(textstem)
library(pairheatmap)
```

```
## Loading required package: grid
```

```
library(LDAvis)

#For successful knitting of document in pdf:
#tinytex::install_tinytex()
```

```
#read data set Tweets May 16, 2020: Covid related hastags as per project document.
getwd()
```

```
## [1] "C:/Ryerson University - Capstone project/Module 2/EIEEE - Large dataset/Combined"
```

```r
data_set_may <- read.csv("corona_tweets_59 May 2020", header = T, sep = ",")
```

```
## Warning in scan(file = file, what = what, sep = sep, quote = quote, dec = dec, :
## embedded nul(s) found in input
```

```r
#take a sample of 1,000, set seed to replicate results across several analysis of methods:
set.seed(1000)
rawData <- data_set_may[sample(nrow(data_set_may), size = 1000), ]
#str(rawData)
```

```r
#Add column id to line up with lab script and transform file to format appropriate for corpus
rawData$id <- 1:nrow(rawData)
#str(rawData)
```

```r
# replace header "text" to "transcript" to line up with lab script

rawData$transcript <- rawData$text
#str(rawData)
```

```r
numberofDocs <- length(rawData$id)
rawData$id <- paste0("Doc", c(1:numberofDocs))
#str(rawData)
```

```r
#set of Corpus using VectorSource() and VCorpus
listofDocs <- tm::VectorSource(rawData$transcript)
listofDocs$Names <- names(rawData$id)
corporaData <- tm::VCorpus(listofDocs)

#use VCorpus as it allows for customized tokenization required for n-gram analysis later on in the code
```

```r
#Lemmatization is the process of reducing a word to its base form while incorporating information about

#Utilizing Thesaurus: lexicon
for(i in 1:1000)
{
    corporaData[[i]]$content <-
    textstem::lemmatize_strings(corporaData[[i]]$content,
                                dictionary = lexicon::hash_lemmas)
}

#Stemming removes a word's suffix (ending), such as es, s, ing, ed, y, based on an heuristic algorithm.

corporaData <- tm::tm_map(corporaData, stemDocument)

#COULDN'T RESOLVE KNITTING ERROR: error in match.fun(FUN) : object 'stemdocument' not found

#remove words that don't add to context of Tweet, but more so are terms that don't distinguish well bet
#Stopword Removal

corporaData <- tm::tm_map(corporaData, removeWords, stopwords('english'))
corporaData <- tm::tm_map(corporaData, removeWords, stopwords('SMART'))
```

```
## Warning: 'stopwords(language = "SMART")' is deprecated.
## Use 'stopwords(source = "smart")' instead.
## See help("Deprecated")
```

```
#Other Pre-processing Steps: Punctuation Marks, Extra Whitespaces, etc
corporaData <- tm::tm_map(corporaData, content_transformer(tolower))
corporaData <- tm::tm_map(corporaData, removePunctuation,
                      ucp = TRUE,
                      preserve_intra_word_contractions = FALSE,
                      preserve_intra_word_dashes = FALSE)
corporaData <- tm::tm_map(corporaData, removeNumbers)
corporaData <- tm::tm_map(corporaData, stripWhitespace)

corporaData[[1]]$content
```

```
## [1] "mani american prove dilig stay home limit spread covid but accept social distanc wane https diq
```

```
# Create a uni-gram Term Document Matrix, #output shows terms for document 1 and frequency in other doc

#output doesn't look right to me, there is something not right with the preprocessed corpus
term.doc.matrix.1g <- tm::TermDocumentMatrix(corporaData)
tm::inspect(term.doc.matrix.1g[1:10,1:10])
```

```
## <<TermDocumentMatrix (terms: 10, documents: 10)>>
## Non-/sparse entries: 0/100
## Sparsity           : 100%
## Maximal term length: 9
## Weighting          : term frequency (tf)
## Sample             :
##                  Docs
## Terms            1 2 3 4 5 6 7 8 9 10
##   $bscuney       0 0 0 0 0 0 0 0 0  0
##   ¦ðÿ            0 0 0 0 0 0 0 0 0  0
##   ¥workplac      0 0 0 0 0 0 0 0 0  0
##   \200ðÿ\230           0 0 0 0 0 0 0 0 0  0
##   +cruis         0 0 0 0 0 0 0 0 0  0
##   +ve            0 0 0 0 0 0 0 0 0  0
##   ¬ðÿ            0 0 0 0 0 0 0 0 0  0
##   ®ðÿ            0 0 0 0 0 0 0 0 0  0
##   \217ðÿ            0 0 0 0 0 0 0 0 0  0
##   \215ðÿ\217           0 0 0 0 0 0 0 0 0  0
```

```
# Represent TDM in a matrix format and display its dimensions. Output shows for each term in every doc,
term.doc.matrix.unigram <- as.matrix(term.doc.matrix.1g)
dim(term.doc.matrix.unigram)
```

```
## [1] 5569 1000
```

```
#head(term.doc.matrix.unigram)
```

```r
# Create a bi-gram Term Document Matrix
tokenizer <- function(x) RWeka::NGramTokenizer(x, RWeka::Weka_control(min=2, max=2))
term.doc.matrix.2g <- tm::TermDocumentMatrix(corporaData, control = list(tokenize=tokenizer))
tm::inspect(term.doc.matrix.2g[1:10,1:10])
```

```
## <<TermDocumentMatrix (terms: 10, documents: 10)>>
## Non-/sparse entries: 0/100
## Sparsity           : 100%
## Maximal term length: 13
## Weighting          : term frequency (tf)
## Sample             :
##                  Docs
## Terms             1 2 3 4 5 6 7 8 9 10
##   $ broadway      0 0 0 0 0 0 0 0 0  0
##   $ child         0 0 0 0 0 0 0 0 0  0
##   $ covid         0 0 0 0 0 0 0 0 0  0
##   $ feder         0 0 0 0 0 0 0 0 0  0
##   $ household     0 0 0 0 0 0 0 0 0  0
##   $ k             0 0 0 0 0 0 0 0 0  0
##   $ million       0 0 0 0 0 0 0 0 0  0
##   $ reptil        0 0 0 0 0 0 0 0 0  0
##   $ trillion      0 0 0 0 0 0 0 0 0  0
##   $bscuney lose 0 0 0 0 0 0 0 0 0  0
```

```r
# Represent TDM in a matrix format and display its dimensions
term.doc.matrix.bigram <- as.matrix(term.doc.matrix.2g)
dim(term.doc.matrix.bigram)
```

```
## [1] 13872  1000
```

```r
#head(term.doc.matrix.bigram)
```

```r
#Getting error: subscript out of bounds?? results in zero terms out of 1000 documents

# Reduce the dimension of the TDM uni-gram matrix
term.doc.matrix.1g <- tm::removeSparseTerms(term.doc.matrix.1g, 0.8)
#tm::inspect(term.doc.matrix.1g[1:10,1:10])

# Represent the TDM as a regular matrix
#term.doc.matrix.unigram <- as.matrix(term.doc.matrix.1g)
#dim(term.doc.matrix.unigram)
#head(term.doc.matrix.unigram)
```

```r
#Getting error: subscript out of bounds?? results in zero terms out of 1000 documents

# Reduce the dimension of the TDM bi-gram matrix
term.doc.matrix.2g <- tm::removeSparseTerms(term.doc.matrix.2g, 0.8)
#tm::inspect(term.doc.matrix.2g[1:10,1:10])

# Represent the TDM as a regular matrix
#term.doc.matrix.bigram <- as.matrix(term.doc.matrix.2g)
#dim(term.doc.matrix.bigram)
#head(term.doc.matrix.bigram)
```

```
#Normalization
# Declaring weights (TF-IDF variants)
tf.idf.weights <- function(tf.vec) {
# Computes tfidf weights from term frequency vector
  n.docs <- length(tf.vec)
  doc.frequency <- length(tf.vec[tf.vec > 0])
  weights <- rep(0, length(tf.vec))
  relative.frequency <- tf.vec[tf.vec > 0] / sum(tf.vec[tf.vec > 0])
  weights[tf.vec > 0] <- relative.frequency *
  log10(1 + n.docs/doc.frequency)
  return(weights)
}
```

```
#Compute the TF-IDF (unigram)
tfidf.matrix.uni <- t(apply(as.matrix(term.doc.matrix.unigram), 1,
FUN = function(row) {tf.idf.weights(row)}))

colnames(tfidf.matrix.uni) <- rawData$id
#head(tfidf.matrix.uni)
dim(tfidf.matrix.uni)
```

```
## [1] 5569 1000
```

```
#Compute the TF-IDF (bigram)
tfidf.matrix.bi <- t(apply(as.matrix(term.doc.matrix.bigram), 1,
FUN = function(row) {tf.idf.weights(row)}))
colnames(tfidf.matrix.bi) <- rawData$id
#head(tfidf.matrix.bi)
dim(tfidf.matrix.bi)
```

```
## [1] 13872  1000
```

```
# index ranges 0 to 1 where 1 means exactly the same and lesser values indicate high, intermediate or l
#Compute Cosine Similarity indices for the uni-gram TDM
c.similarity.matrix.uni <-
text2vec::sim2(t(tfidf.matrix.uni), method = 'cosine')
```

```
#Compute Cosine Similarity Indices for the bi-gram TDM
c.similarity.matrix.bi <-
text2vec::sim2(t(tfidf.matrix.bi), method = 'cosine')
```

```
#Display Ranked Lists for last tweet in sample
sort(c.similarity.matrix.uni[1000, ], decreasing = TRUE)[1:1000]
```

```
##      Doc1000        Doc66       Doc249       Doc863       Doc521       Doc852 
## 1.000000e+00 5.514006e-02 6.580743e-03 6.542373e-03 4.314750e-03 4.239623e-03 
##       Doc375       Doc613       Doc204       Doc319       Doc827       Doc254 
## 3.669922e-03 3.001138e-03 2.258314e-03 6.726797e-04 6.726797e-04 4.353545e-04 
##       Doc517       Doc924       Doc310       Doc990       Doc548       Doc605 
## 4.353545e-04 4.353545e-04 4.122757e-04 4.122757e-04 3.762658e-04 3.473827e-04 
##       Doc242         Doc1       Doc826       Doc149       Doc250       Doc231 
## 2.956028e-04 2.905642e-04 2.831963e-04 2.562009e-04 2.402550e-04 2.338094e-04
##       Doc859       Doc640       Doc112       Doc460       Doc718       Doc472
## 2.280266e-04 2.262618e-04 2.165692e-04 2.129159e-04 2.127279e-04 2.113713e-04
##       Doc282       Doc358       Doc803       Doc809       Doc851       Doc393
## 1.987172e-04 1.951749e-04 1.777768e-04 1.735612e-04 1.647821e-04 1.573340e-04
##       Doc973       Doc464       Doc714       Doc931       Doc227       Doc473
## 1.573340e-04 1.416437e-04 1.402924e-04 1.372452e-04 1.346385e-04 1.346385e-04
##       Doc615       Doc710       Doc874       Doc984       Doc505       Doc791
## 1.346385e-04 1.346385e-04 1.346385e-04 1.346385e-04 1.314499e-04 1.230784e-04
##       Doc523       Doc269       Doc900       Doc538       Doc736        Doc68
## 1.177220e-04 1.140213e-04 1.059971e-04 1.000675e-04 9.570863e-05 9.531941e-05
##        Doc24       Doc992        Doc74       Doc105       Doc961       Doc547
## 8.857525e-05 8.691518e-05 8.261753e-05 8.216124e-05 7.731465e-05 7.530749e-05
##       Doc688       Doc864       Doc675       Doc782         Doc7       Doc167
## 7.483624e-05 6.982542e-05 6.787843e-05 6.467287e-05 6.331333e-05 6.051159e-05
##       Doc823        Doc28       Doc299       Doc963       Doc271       Doc715
## 5.854631e-05 5.598389e-05 5.598389e-05 5.264303e-05 4.355356e-05 4.355356e-05
##         Doc6       Doc866       Doc291       Doc155       Doc491        Doc61
## 4.256291e-05 4.256291e-05 4.017079e-05 3.940318e-05 3.922173e-05 3.891908e-05
##       Doc138       Doc408       Doc434       Doc690       Doc506       Doc388
## 3.891908e-05 3.799845e-05 3.673523e-05 3.328182e-05 3.272983e-05 3.107495e-05
##       Doc948       Doc438        Doc27       Doc376       Doc597       Doc965
## 2.723674e-05 2.626651e-05 2.571079e-05 2.332888e-05 2.274006e-05 2.248784e-05
##       Doc775        Doc21       Doc368       Doc719       Doc729       Doc188
## 2.231511e-05 2.220811e-05 2.206437e-05 2.206226e-05 2.107747e-05 1.990681e-05
##       Doc646       Doc993       Doc530        Doc50        Doc72       Doc365
## 1.949942e-05 1.946104e-05 1.887686e-05 1.878699e-05 1.875653e-05 1.859861e-05
##       Doc622       Doc746       Doc551        Doc85       Doc749       Doc943
## 1.805497e-05 1.751705e-05 1.731436e-05 1.712519e-05 1.704962e-05 1.702386e-05
##       Doc629       Doc850       Doc370        Doc46       Doc892       Doc957
## 1.646842e-05 1.550898e-05 1.526128e-05 1.502580e-05 1.502455e-05 1.460340e-05
##       Doc699       Doc936        Doc75       Doc379        Doc55       Doc423
## 1.433207e-05 1.420610e-05 1.261430e-05 1.201417e-05 4.866673e-08 4.866673e-08
##       Doc482       Doc890       Doc336       Doc511       Doc262       Doc273
## 4.866673e-08 4.866673e-08 4.254761e-08 4.254761e-08 3.970097e-08 3.970097e-08
##       Doc983       Doc999       Doc135       Doc490       Doc595        Doc30
## 3.970097e-08 3.970097e-08 3.606198e-08 3.606198e-08 3.606198e-08 3.444966e-08
##       Doc202       Doc157       Doc187       Doc614       Doc175       Doc192
## 3.444966e-08 3.011157e-08 3.011157e-08 3.011157e-08 2.875647e-08 2.875647e-08
##       Doc875       Doc561       Doc808        Doc84       Doc395       Doc583
## 2.875647e-08 2.838417e-08 2.838417e-08 2.816981e-08 2.816981e-08 2.667705e-08
##       Doc972        Doc82       Doc470       Doc998       Doc326       Doc327
## 2.667705e-08 2.591477e-08 2.399810e-08 2.399187e-08 2.388393e-08 2.305832e-08
##       Doc573       Doc618       Doc248       Doc363       Doc111       Doc329
## 2.305300e-08 2.305300e-08 2.299289e-08 2.287413e-08 2.264351e-08 2.227547e-08
##       Doc119       Doc445        Doc97       Doc762       Doc466       Doc836
## 2.145106e-08 2.145106e-08 2.099629e-08 2.099629e-08 2.086613e-08 2.070760e-08
##        Doc87       Doc520       Doc114       Doc133       Doc770       Doc790
## 2.040477e-08 2.040477e-08 2.026123e-08 2.026123e-08 2.012502e-08 2.012502e-08
##       Doc593       Doc598       Doc516       Doc456       Doc311       Doc401
## 1.959636e-08 1.957277e-08 1.948421e-08 1.947129e-08 1.934270e-08 1.926154e-08
##       Doc911       Doc539       Doc533       Doc446       Doc802       Doc895
## 1.925688e-08 1.924586e-08 1.916107e-08 1.864065e-08 1.843927e-08 1.832374e-08
##       Doc601       Doc278       Doc346       Doc467       Doc660       Doc382
## 1.825313e-08 1.823954e-08 1.823954e-08 1.754153e-08 1.754153e-08 1.746950e-08
##       Doc448       Doc592       Doc687       Doc986       Doc700        Doc40
## 1.746950e-08 1.743016e-08 1.725427e-08 1.697369e-08 1.696902e-08 1.694516e-08
##       Doc141       Doc380        Doc60       Doc296       Doc801       Doc206
## 1.693940e-08 1.693595e-08 1.693404e-08 1.691810e-08 1.690908e-08 1.690292e-08
##       Doc268       Doc970       Doc469       Doc873       Doc140       Doc403
## 1.689663e-08 1.688656e-08 1.686939e-08 1.686191e-08 1.680047e-08 1.679730e-08
##       Doc433       Doc142       Doc794       Doc174       Doc903       Doc260
## 1.678510e-08 1.676713e-08 1.676210e-08 1.675886e-08 1.674278e-08 1.672235e-08
##       Doc716       Doc181       Doc410       Doc544       Doc784       Doc627
## 1.669128e-08 1.667211e-08 1.665124e-08 1.663778e-08 1.661669e-08 1.660526e-08
##       Doc284       Doc136       Doc301       Doc462       Doc178       Doc686
## 1.658708e-08 1.655602e-08 1.655333e-08 1.653704e-08 1.652325e-08 1.649500e-08
##       Doc304       Doc879       Doc128       Doc125       Doc361        Doc36
## 1.648790e-08 1.648464e-08 1.640790e-08 1.639808e-08 1.638696e-08 1.638013e-08
##       Doc704       Doc843        Doc48       Doc137       Doc910        Doc64
## 1.636856e-08 1.636856e-08 1.636331e-08 1.634367e-08 1.634367e-08 1.631325e-08
##       Doc265       Doc789       Doc429       Doc946       Doc124       Doc146
## 1.629705e-08 1.622732e-08 1.622200e-08 1.617447e-08 1.608670e-08 1.607461e-08
##       Doc351       Doc654        Doc23       Doc152       Doc787       Doc235
## 1.605413e-08 1.604201e-08 1.603274e-08 1.600189e-08 1.586891e-08 1.585757e-08
##        Doc42       Doc148        Doc22       Doc207       Doc735       Doc500
## 1.575528e-08 1.571925e-08 1.569076e-08 1.552050e-08 1.552050e-08 1.547666e-08
##        Doc47       Doc528       Doc761       Doc953       Doc588       Doc845
## 1.545882e-08 1.538984e-08 1.538013e-08 1.537468e-08 1.532429e-08 1.531134e-08
##       Doc741       Doc171       Doc976       Doc846       Doc453       Doc907
## 1.526017e-08 1.522813e-08 1.522548e-08 1.516492e-08 1.514826e-08 1.505061e-08
##       Doc712        Doc92       Doc996       Doc844       Doc860       Doc865
## 1.501684e-08 1.498464e-08 1.498464e-08 1.497721e-08 1.485249e-08 1.476782e-08
##       Doc647       Doc163       Doc638       Doc898       Doc555       Doc459
## 1.474699e-08 1.474311e-08 1.458456e-08 1.443607e-08 1.443126e-08 1.432069e-08
##       Doc723       Doc303       Doc236       Doc535       Doc384       Doc776
## 1.429538e-08 1.427373e-08 1.426274e-08 1.423720e-08 1.421218e-08 1.397335e-08
##       Doc725       Doc881       Doc842       Doc847       Doc913       Doc334
## 1.388227e-08 1.385896e-08 1.368790e-08 1.364422e-08 1.360424e-08 1.356294e-08
##       Doc822       Doc960       Doc989       Doc131       Doc251        Doc31
## 1.342274e-08 1.342274e-08 1.326413e-08 1.316892e-08 1.310705e-08 1.273233e-08
##       Doc357       Doc584        Doc51       Doc220       Doc867       Doc150
## 1.272029e-08 1.264465e-08 1.263034e-08 1.260215e-08 1.260215e-08 1.256557e-08
##       Doc546       Doc397       Doc425        Doc98       Doc377       Doc447
## 1.253089e-08 1.249514e-08 1.249514e-08 1.247068e-08 1.236659e-08 1.211012e-08
##       Doc504       Doc878       Doc814        Doc13       Doc121       Doc758
## 1.200221e-08 1.200212e-08 1.200200e-08 1.200119e-08 1.199897e-08 1.199819e-08
##       Doc367       Doc416       Doc708       Doc412        Doc90       Doc126
## 1.199591e-08 1.198803e-08 1.197660e-08 1.196856e-08 1.195941e-08 1.194960e-08
##       Doc199       Doc577       Doc623       Doc427       Doc461       Doc279
## 1.194034e-08 1.193891e-08 1.193708e-08 1.191945e-08 1.191187e-08 1.190122e-08
##       Doc239       Doc738       Doc695       Doc587       Doc793       Doc233
## 1.188587e-08 1.188247e-08 1.188088e-08 1.187165e-08 1.186169e-08 1.185361e-08
##       Doc132       Doc920        Doc52       Doc958       Doc247       Doc474
## 1.184302e-08 1.183466e-08 1.182032e-08 1.181819e-08 1.180868e-08 1.177300e-08
##       Doc835       Doc904       Doc147       Doc971       Doc320       Doc934
## 1.177300e-08 1.177300e-08 1.177299e-08 1.177260e-08 1.177199e-08 1.176887e-08
##       Doc176       Doc522       Doc151       Doc287       Doc120       Doc394
## 1.176788e-08 1.176172e-08 1.176029e-08 1.175935e-08 1.175686e-08 1.175616e-08
##       Doc309       Doc109       Doc652       Doc255       Doc857       Doc887
## 1.174321e-08 1.170327e-08 1.169839e-08 1.168190e-08 1.166139e-08 1.161551e-08
##       Doc305       Doc537       Doc399       Doc978       Doc632       Doc369
## 1.157939e-08 1.157024e-08 1.153835e-08 1.153627e-08 1.146697e-08 1.143630e-08
##       Doc253        Doc67        Doc16       Doc435       Doc938       Doc292
## 1.143629e-08 1.142951e-08 1.142567e-08 1.141858e-08 1.141653e-08 1.141435e-08
##        Doc76       Doc649        Doc94       Doc314       Doc930       Doc297
## 1.141349e-08 1.140882e-08 1.140502e-08 1.139133e-08 1.138015e-08 1.137738e-08
##         Doc4         Doc9       Doc684       Doc839       Doc568       Doc893
## 1.135615e-08 1.124292e-08 1.124041e-08 1.123850e-08 1.123724e-08 1.123365e-08
##       Doc981       Doc552       Doc921       Doc103       Doc217       Doc290
## 1.122626e-08 1.122289e-08 1.120839e-08 1.120383e-08 1.115689e-08 1.113189e-08
##       Doc928       Doc668       Doc485       Doc643       Doc288        Doc56
## 1.112712e-08 1.112469e-08 1.111149e-08 1.110473e-08 1.109238e-08 1.106323e-08
##       Doc374       Doc527       Doc422       Doc681       Doc856        Doc79
## 1.098300e-08 1.098300e-08 1.095885e-08 1.094545e-08 1.085868e-08 1.080996e-08
##       Doc607       Doc819       Doc869       Doc183       Doc685       Doc955
## 1.079801e-08 1.075553e-08 1.075363e-08 1.070451e-08 1.070451e-08 1.062626e-08
##       Doc633       Doc733       Doc428       Doc915       Doc565       Doc731
## 1.058768e-08 1.049073e-08 1.048798e-08 1.047722e-08 1.045233e-08 1.043015e-08
##       Doc229       Doc525       Doc580       Doc407       Doc882       Doc558
## 1.033475e-08 1.029185e-08 1.023185e-08 1.019937e-08 1.014866e-08 1.008159e-08
##       Doc917       Doc967       Doc340        Doc57       Doc277       Doc648
## 1.004648e-08 1.001380e-08 1.001044e-08 9.984604e-09 9.959582e-09 9.889735e-09
##       Doc959       Doc628        Doc96       Doc611       Doc182       Doc345
## 9.822986e-09 9.799767e-09 9.799636e-09 9.793331e-09 9.782932e-09 9.781497e-09
##       Doc439       Doc721       Doc724       Doc625        Doc45       Doc195
## 9.781079e-09 9.780546e-09 9.777925e-09 9.776664e-09 9.752262e-09 9.741004e-09
##       Doc871       Doc824       Doc626        Doc91       Doc812       Doc512
## 9.734361e-09 9.728304e-09 9.719079e-09 9.712013e-09 9.705643e-09 9.702789e-09
##       Doc441       Doc594       Doc902       Doc272         Doc5       Doc759
## 9.702650e-09 9.702459e-09 9.696829e-09 9.694552e-09 9.681520e-09 9.671202e-09
##       Doc503       Doc554       Doc526       Doc661       Doc360       Doc641
## 9.667728e-09 9.639709e-09 9.630195e-09 9.624566e-09 9.609395e-09 9.607522e-09
##       Doc799       Doc339        Doc10       Doc417       Doc113        Doc89
## 9.602448e-09 9.582482e-09 9.539766e-09 9.529857e-09 9.518389e-09 9.515605e-09
##       Doc364       Doc559       Doc295       Doc586       Doc894       Doc252
## 9.513527e-09 9.486886e-09 9.481982e-09 9.473683e-09 9.457539e-09 9.447637e-09
##       Doc667        Doc12       Doc912       Doc639       Doc987       Doc532
## 9.447286e-09 9.445530e-09 9.439651e-09 9.431580e-09 9.430131e-09 9.427441e-09
##       Doc707        Doc59       Doc755       Doc283       Doc840        Doc37
## 9.422805e-09 9.422598e-09 9.410492e-09 9.406958e-09 9.404491e-09 9.402503e-09
##       Doc463       Doc106       Doc419       Doc349       Doc945       Doc574
## 9.396285e-09 9.378320e-09 9.358360e-09 9.351093e-09 9.324236e-09 9.318522e-09
##       Doc286        Doc58       Doc222       Doc356       Doc610       Doc829
## 9.306410e-09 9.296570e-09 9.292478e-09 9.278009e-09 9.203367e-09 9.198741e-09
##       Doc834         Doc2        Doc18       Doc671       Doc209       Doc238
## 9.195958e-09 9.171265e-09 9.165644e-09 9.128766e-09 9.126385e-09 9.104129e-09
##       Doc190       Doc655       Doc196       Doc929       Doc906       Doc519
## 9.016364e-09 9.015731e-09 8.961707e-09 8.958197e-09 8.926634e-09 8.924464e-09
##       Doc221       Doc509       Doc969        Doc33       Doc342       Doc420
## 8.894680e-09 8.888331e-09 8.806263e-09 8.588207e-09 8.488775e-09 8.486847e-09
##       Doc682       Doc241       Doc518       Doc331       Doc449       Doc997
## 8.486847e-09 8.485960e-09 8.483694e-09 8.483202e-09 8.466521e-09 8.461280e-09
##       Doc328       Doc164       Doc680       Doc780       Doc390       Doc672
## 8.460077e-09 8.444231e-09 8.439564e-09 8.433740e-09 8.428479e-09 8.420819e-09
##       Doc330       Doc110       Doc381       Doc502       Doc868       Doc219
## 8.418552e-09 8.418226e-09 8.417501e-09 8.414843e-09 8.401332e-09 8.393895e-09
##       Doc536       Doc748       Doc127       Doc529       Doc940       Doc754
## 8.391277e-09 8.372186e-09 8.358235e-09 8.308040e-09 8.298671e-09 8.283874e-09
##       Doc608       Doc352       Doc820       Doc487       Doc952       Doc620
## 8.279932e-09 8.279035e-09 8.278474e-09 8.276530e-09 8.260356e-09 8.256349e-09
##       Doc230       Doc925       Doc116       Doc210       Doc677       Doc631
## 8.254830e-09 8.237684e-09 8.217492e-09 8.208560e-09 8.205306e-09 8.204753e-09
##       Doc191       Doc389       Doc436       Doc786       Doc421       Doc115
## 8.202662e-09 8.199927e-09 8.196146e-09 8.193607e-09 8.193323e-09 8.188956e-09
##       Doc730       Doc185       Doc788       Doc471       Doc567       Doc285
## 8.170640e-09 8.169480e-09 8.165535e-09 8.165497e-09 8.154880e-09 8.151741e-09
##        Doc19       Doc914       Doc205       Doc624         Doc3       Doc437
## 8.134175e-09 8.127020e-09 8.119850e-09 8.105958e-09 8.100070e-09 8.097825e-09
##       Doc926       Doc225       Doc256       Doc602       Doc406       Doc662
## 8.087449e-09 8.077545e-09 8.075419e-09 8.074233e-09 8.066952e-09 8.062236e-09
##       Doc763       Doc158       Doc348       Doc244        Doc73       Doc747
## 8.000441e-09 7.944848e-09 7.935405e-09 7.932144e-09 7.915859e-09 7.909389e-09
##       Doc732       Doc414       Doc696       Doc585       Doc338       Doc237
## 7.888345e-09 7.826430e-09 7.766088e-09 7.752593e-09 7.676197e-09 7.642184e-09
##        Doc80       Doc440       Doc901       Doc698       Doc889        Doc63
## 7.636569e-09 7.610802e-09 7.590867e-09 7.590860e-09 7.589075e-09 7.581379e-09
##       Doc267       Doc270       Doc347       Doc201       Doc494       Doc722
## 7.578511e-09 7.559241e-09 7.559213e-09 7.555489e-09 7.535309e-09 7.530937e-09
##       Doc450       Doc691       Doc769       Doc604       Doc923       Doc415
## 7.527601e-09 7.525220e-09 7.513163e-09 7.509756e-09 7.508150e-09 7.506577e-09
##        Doc34       Doc665       Doc701        Doc17       Doc919       Doc876
## 7.490600e-09 7.465640e-09 7.465604e-09 7.440245e-09 7.431694e-09 7.423728e-09
##       Doc468       Doc452       Doc117       Doc933       Doc337       Doc816
## 7.422677e-09 7.418713e-09 7.413429e-09 7.412715e-09 7.403500e-09 7.400861e-09
##       Doc837       Doc600       Doc757       Doc405       Doc727       Doc908
## 7.400461e-09 7.400394e-09 7.394635e-09 7.388701e-09 7.373180e-09 7.369647e-09
##       Doc697       Doc941       Doc316       Doc108       Doc774       Doc180
## 7.363728e-09 7.348495e-09 7.330389e-09 7.325003e-09 7.323276e-09 7.318827e-09
##       Doc160       Doc557       Doc553       Doc918       Doc768       Doc571
## 7.314218e-09 7.301465e-09 7.298830e-09 7.294297e-09 7.286816e-09 7.285165e-09
##       Doc524       Doc398       Doc514       Doc161       Doc673        Doc78
## 7.271070e-09 7.242921e-09 7.235079e-09 7.223207e-09 7.205244e-09 7.202793e-09
##       Doc950        Doc71       Doc477       Doc302       Doc431       Doc315
## 7.195636e-09 7.185661e-09 7.155089e-09 7.154873e-09 7.124904e-09 7.104309e-09
##       Doc481        Doc14       Doc797       Doc413       Doc982        Doc86
## 7.083888e-09 7.064461e-09 7.060944e-09 7.019642e-09 7.003870e-09 7.001947e-09
##       Doc683       Doc550       Doc211       Doc411       Doc308       Doc497
## 6.993383e-09 6.973194e-09 6.972332e-09 6.970555e-09 6.966331e-09 6.939921e-09
##       Doc501       Doc122       Doc275       Doc877       Doc324       Doc988
## 6.929277e-09 6.899138e-09 6.886708e-09 6.878808e-09 6.873080e-09 6.853923e-09
##       Doc853        Doc44       Doc566       Doc658       Doc443       Doc678
## 6.839558e-09 6.833182e-09 6.821360e-09 6.817816e-09 6.814833e-09 6.814333e-09
##       Doc353       Doc642       Doc951        Doc69       Doc335        Doc38
## 6.780925e-09 6.777675e-09 6.772186e-09 6.770727e-09 6.760188e-09 6.757780e-09
##       Doc457       Doc100       Doc603       Doc619       Doc179       Doc354
## 6.742825e-09 6.742483e-09 6.736380e-09 6.719073e-09 6.710106e-09 6.689274e-09
##       Doc234       Doc581       Doc693       Doc713       Doc612       Doc991
## 6.685742e-09 6.682901e-09 6.680769e-09 6.677413e-09 6.663341e-09 6.626467e-09
##       Doc591       Doc760       Doc246       Doc711       Doc767       Doc706
## 6.602772e-09 6.600990e-09 6.595578e-09 6.584106e-09 6.567368e-09 6.541908e-09
##       Doc424       Doc564       Doc162       Doc792       Doc644       Doc956
## 6.528883e-09 6.518913e-09 6.505381e-09 6.491207e-09 6.404114e-09 6.387367e-09
##       Doc245       Doc954       Doc854       Doc513       Doc651       Doc645
## 6.361779e-09 6.354403e-09 6.350750e-09 6.332751e-09 6.330808e-09 6.321604e-09
##       Doc772       Doc947       Doc280       Doc773       Doc888       Doc811
## 6.318856e-09 6.312535e-09 6.304713e-09 6.303989e-09 6.280961e-09 6.272379e-09
##       Doc476       Doc609       Doc478       Doc534        Doc20       Doc359
## 6.266696e-09 6.264057e-09 6.262454e-09 6.254821e-09 6.226109e-09 6.185348e-09
##       Doc556       Doc905       Doc545       Doc709       Doc570       Doc702
## 6.169299e-09 6.162263e-09 6.161080e-09 6.161043e-09 6.154392e-09 6.146744e-09
##       Doc831       Doc232       Doc444       Doc343        Doc53       Doc752
## 6.143879e-09 6.120765e-09 6.110927e-09 6.030632e-09 6.022063e-09 5.967091e-09
##       Doc985       Doc484       Doc200       Doc243       Doc169       Doc274
## 5.957635e-09 5.940795e-09 5.929034e-09 5.920188e-09 5.904226e-09 5.876847e-09
##       Doc884       Doc259       Doc785       Doc289       Doc962        Doc88
## 5.870447e-09 5.858074e-09 5.857876e-09 5.841953e-09 5.834490e-09 5.826829e-09
##       Doc495       Doc475       Doc306       Doc750       Doc392        Doc43
## 5.814740e-09 5.798489e-09 5.768880e-09 5.751382e-09 5.739025e-09 5.729920e-09
##       Doc541        Doc81         Doc8       Doc156       Doc224       Doc153
## 5.715576e-09 5.656179e-09 5.629715e-09 5.599978e-09 5.571510e-09 5.561346e-09
##       Doc173       Doc781       Doc194       Doc935       Doc742       Doc215
## 5.557036e-09 5.539238e-09 5.521099e-09 5.516410e-09 5.506259e-09 5.490976e-09
##       Doc637       Doc798       Doc897       Doc968       Doc213        Doc93
## 5.487562e-09 5.463003e-09 5.460419e-09 5.459522e-09 5.426645e-09 5.420051e-09
##       Doc184       Doc616       Doc430       Doc806       Doc896       Doc664
## 5.387851e-09 5.361985e-09 5.310575e-09 5.303593e-09 5.297188e-09 5.283309e-09
##       Doc821       Doc575       Doc692       Doc635       Doc861       Doc409
## 5.275099e-09 5.269876e-09 5.257828e-09 5.244322e-09 5.236606e-09 5.209476e-09
##       Doc228       Doc129       Doc294       Doc805       Doc226       Doc323
## 5.174962e-09 5.153244e-09 5.009616e-09 5.003775e-09 4.980965e-09 4.974135e-09
##       Doc606       Doc634       Doc670        Doc62       Doc582        Doc49
## 4.953629e-09 4.871290e-09 4.854624e-09 4.806299e-09 4.786432e-09 4.777023e-09
##       Doc872       Doc771       Doc744       Doc216       Doc765       Doc107
## 4.739977e-09 4.699087e-09 4.673271e-09 4.613055e-09 4.526347e-09 4.503761e-09
##       Doc630       Doc994       Doc293       Doc313       Doc510       Doc166
## 4.502589e-09 4.404591e-09 4.323807e-09 4.236564e-09 4.100254e-09 3.964306e-09
##       Doc426       Doc159        Doc11        Doc15        Doc25        Doc26
## 3.871077e-09 3.411644e-09 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##        Doc29        Doc32        Doc35        Doc39        Doc41        Doc54
## 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##        Doc65        Doc70        Doc77        Doc83        Doc95        Doc99
## 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##       Doc101       Doc102       Doc104       Doc118       Doc123       Doc130
## 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##       Doc134       Doc139       Doc143       Doc144       Doc145       Doc154
## 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##       Doc165       Doc168       Doc170       Doc172       Doc177       Doc186
## 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##       Doc189       Doc193       Doc197       Doc198       Doc203       Doc208
## 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##       Doc212       Doc214       Doc218       Doc223       Doc240       Doc257
## 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##       Doc258       Doc261       Doc263       Doc264       Doc266       Doc276
## 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##       Doc281       Doc298       Doc300       Doc307       Doc312       Doc317
## 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##       Doc318       Doc321       Doc322       Doc325       Doc332       Doc333
## 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##       Doc341       Doc344       Doc350       Doc355       Doc362       Doc366
## 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##       Doc371       Doc372       Doc373       Doc378       Doc383       Doc385
## 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##       Doc386       Doc387       Doc391       Doc396       Doc400       Doc402
## 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##       Doc404       Doc418       Doc432       Doc442       Doc451       Doc454
## 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##       Doc455       Doc458       Doc465       Doc479       Doc480       Doc483
## 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##       Doc486       Doc488       Doc489       Doc492       Doc493       Doc496
## 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##       Doc498       Doc499       Doc507       Doc508       Doc515       Doc531
## 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##       Doc540       Doc542       Doc543       Doc549       Doc560       Doc562
## 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##       Doc563       Doc569       Doc572       Doc576       Doc578       Doc579
## 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##       Doc589       Doc590       Doc596       Doc599       Doc617       Doc621
## 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##       Doc636       Doc650       Doc653       Doc656       Doc657       Doc659
## 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##       Doc663       Doc666       Doc669       Doc674       Doc676       Doc679
## 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##       Doc689       Doc694       Doc703       Doc705       Doc717       Doc720
## 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##       Doc726       Doc728       Doc734       Doc737       Doc739       Doc740
## 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##       Doc743       Doc745       Doc751       Doc753       Doc756       Doc764
## 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##       Doc766       Doc777       Doc778       Doc779       Doc783       Doc795
## 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##       Doc796       Doc800       Doc804       Doc807       Doc810       Doc813
## 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##       Doc815       Doc817       Doc818       Doc825       Doc828       Doc830
## 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##       Doc832       Doc833       Doc838       Doc841       Doc848       Doc849
## 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##       Doc855       Doc858       Doc862       Doc870       Doc880       Doc883
## 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##       Doc885       Doc886       Doc891       Doc899       Doc909       Doc916
## 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##       Doc922       Doc927       Doc932       Doc937       Doc939       Doc942
## 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##       Doc944       Doc949       Doc964       Doc966       Doc974       Doc975
## 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
##       Doc977       Doc979       Doc980       Doc995
## 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
```

```
sort(c.similarity.matrix.bi[1000, ], decreasing = TRUE)[1:1000]
```

```
##    Doc1000     Doc791       Doc1       Doc2       Doc3       Doc4       Doc5
## 1.00000000 0.03890828 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##       Doc6       Doc7       Doc8       Doc9      Doc10      Doc11      Doc12
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##      Doc13      Doc14      Doc15      Doc16      Doc17      Doc18      Doc19
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##      Doc20      Doc21      Doc22      Doc23      Doc24      Doc25      Doc26
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##      Doc27      Doc28      Doc29      Doc30      Doc31      Doc32      Doc33
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##      Doc34      Doc35      Doc36      Doc37      Doc38      Doc39      Doc40
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##      Doc41      Doc42      Doc43      Doc44      Doc45      Doc46      Doc47
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##      Doc48      Doc49      Doc50      Doc51      Doc52      Doc53      Doc54
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##      Doc55      Doc56      Doc57      Doc58      Doc59      Doc60      Doc61
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##      Doc62      Doc63      Doc64      Doc65      Doc66      Doc67      Doc68
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##      Doc69      Doc70      Doc71      Doc72      Doc73      Doc74      Doc75
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##      Doc76      Doc77      Doc78      Doc79      Doc80      Doc81      Doc82
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##      Doc83      Doc84      Doc85      Doc86      Doc87      Doc88      Doc89
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##      Doc90      Doc91      Doc92      Doc93      Doc94      Doc95      Doc96
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##      Doc97      Doc98      Doc99     Doc100     Doc101     Doc102     Doc103
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc104     Doc105     Doc106     Doc107     Doc108     Doc109     Doc110
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc111     Doc112     Doc113     Doc114     Doc115     Doc116     Doc117
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc118     Doc119     Doc120     Doc121     Doc122     Doc123     Doc124
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc125     Doc126     Doc127     Doc128     Doc129     Doc130     Doc131
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc132     Doc133     Doc134     Doc135     Doc136     Doc137     Doc138
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc139     Doc140     Doc141     Doc142     Doc143     Doc144     Doc145
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc146     Doc147     Doc148     Doc149     Doc150     Doc151     Doc152
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc153     Doc154     Doc155     Doc156     Doc157     Doc158     Doc159
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc160     Doc161     Doc162     Doc163     Doc164     Doc165     Doc166
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc167     Doc168     Doc169     Doc170     Doc171     Doc172     Doc173 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc174     Doc175     Doc176     Doc177     Doc178     Doc179     Doc180 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc181     Doc182     Doc183     Doc184     Doc185     Doc186     Doc187 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc188     Doc189     Doc190     Doc191     Doc192     Doc193     Doc194 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc195     Doc196     Doc197     Doc198     Doc199     Doc200     Doc201 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc202     Doc203     Doc204     Doc205     Doc206     Doc207     Doc208 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc209     Doc210     Doc211     Doc212     Doc213     Doc214     Doc215 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc216     Doc217     Doc218     Doc219     Doc220     Doc221     Doc222 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc223     Doc224     Doc225     Doc226     Doc227     Doc228     Doc229 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc230     Doc231     Doc232     Doc233     Doc234     Doc235     Doc236 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc237     Doc238     Doc239     Doc240     Doc241     Doc242     Doc243 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc244     Doc245     Doc246     Doc247     Doc248     Doc249     Doc250 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc251     Doc252     Doc253     Doc254     Doc255     Doc256     Doc257 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc258     Doc259     Doc260     Doc261     Doc262     Doc263     Doc264 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc265     Doc266     Doc267     Doc268     Doc269     Doc270     Doc271 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc272     Doc273     Doc274     Doc275     Doc276     Doc277     Doc278 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc279     Doc280     Doc281     Doc282     Doc283     Doc284     Doc285 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc286     Doc287     Doc288     Doc289     Doc290     Doc291     Doc292 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc293     Doc294     Doc295     Doc296     Doc297     Doc298     Doc299 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc300     Doc301     Doc302     Doc303     Doc304     Doc305     Doc306 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc307     Doc308     Doc309     Doc310     Doc311     Doc312     Doc313 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc314     Doc315     Doc316     Doc317     Doc318     Doc319     Doc320 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc321     Doc322     Doc323     Doc324     Doc325     Doc326     Doc327 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc328     Doc329     Doc330     Doc331     Doc332     Doc333     Doc334 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc335     Doc336     Doc337     Doc338     Doc339     Doc340     Doc341 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc342     Doc343     Doc344     Doc345     Doc346     Doc347     Doc348 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc349     Doc350     Doc351     Doc352     Doc353     Doc354     Doc355 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc356     Doc357     Doc358     Doc359     Doc360     Doc361     Doc362
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc363     Doc364     Doc365     Doc366     Doc367     Doc368     Doc369
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc370     Doc371     Doc372     Doc373     Doc374     Doc375     Doc376
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc377     Doc378     Doc379     Doc380     Doc381     Doc382     Doc383
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc384     Doc385     Doc386     Doc387     Doc388     Doc389     Doc390
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc391     Doc392     Doc393     Doc394     Doc395     Doc396     Doc397
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc398     Doc399     Doc400     Doc401     Doc402     Doc403     Doc404
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc405     Doc406     Doc407     Doc408     Doc409     Doc410     Doc411
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc412     Doc413     Doc414     Doc415     Doc416     Doc417     Doc418
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc419     Doc420     Doc421     Doc422     Doc423     Doc424     Doc425
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc426     Doc427     Doc428     Doc429     Doc430     Doc431     Doc432
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc433     Doc434     Doc435     Doc436     Doc437     Doc438     Doc439
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc440     Doc441     Doc442     Doc443     Doc444     Doc445     Doc446
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc447     Doc448     Doc449     Doc450     Doc451     Doc452     Doc453
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc454     Doc455     Doc456     Doc457     Doc458     Doc459     Doc460
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc461     Doc462     Doc463     Doc464     Doc465     Doc466     Doc467
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc468     Doc469     Doc470     Doc471     Doc472     Doc473     Doc474
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc475     Doc476     Doc477     Doc478     Doc479     Doc480     Doc481
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc482     Doc483     Doc484     Doc485     Doc486     Doc487     Doc488
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc489     Doc490     Doc491     Doc492     Doc493     Doc494     Doc495
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc496     Doc497     Doc498     Doc499     Doc500     Doc501     Doc502
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc503     Doc504     Doc505     Doc506     Doc507     Doc508     Doc509
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc510     Doc511     Doc512     Doc513     Doc514     Doc515     Doc516
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc517     Doc518     Doc519     Doc520     Doc521     Doc522     Doc523
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc524     Doc525     Doc526     Doc527     Doc528     Doc529     Doc530
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc531     Doc532     Doc533     Doc534     Doc535     Doc536     Doc537
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc538     Doc539     Doc540     Doc541     Doc542     Doc543     Doc544
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc545     Doc546     Doc547     Doc548     Doc549     Doc550     Doc551 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc552     Doc553     Doc554     Doc555     Doc556     Doc557     Doc558 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc559     Doc560     Doc561     Doc562     Doc563     Doc564     Doc565 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc566     Doc567     Doc568     Doc569     Doc570     Doc571     Doc572 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc573     Doc574     Doc575     Doc576     Doc577     Doc578     Doc579 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc580     Doc581     Doc582     Doc583     Doc584     Doc585     Doc586 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc587     Doc588     Doc589     Doc590     Doc591     Doc592     Doc593 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc594     Doc595     Doc596     Doc597     Doc598     Doc599     Doc600 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc601     Doc602     Doc603     Doc604     Doc605     Doc606     Doc607 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc608     Doc609     Doc610     Doc611     Doc612     Doc613     Doc614 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc615     Doc616     Doc617     Doc618     Doc619     Doc620     Doc621 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc622     Doc623     Doc624     Doc625     Doc626     Doc627     Doc628 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc629     Doc630     Doc631     Doc632     Doc633     Doc634     Doc635 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc636     Doc637     Doc638     Doc639     Doc640     Doc641     Doc642 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc643     Doc644     Doc645     Doc646     Doc647     Doc648     Doc649 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc650     Doc651     Doc652     Doc653     Doc654     Doc655     Doc656 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc657     Doc658     Doc659     Doc660     Doc661     Doc662     Doc663 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc664     Doc665     Doc666     Doc667     Doc668     Doc669     Doc670 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc671     Doc672     Doc673     Doc674     Doc675     Doc676     Doc677 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc678     Doc679     Doc680     Doc681     Doc682     Doc683     Doc684 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc685     Doc686     Doc687     Doc688     Doc689     Doc690     Doc691 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc692     Doc693     Doc694     Doc695     Doc696     Doc697     Doc698 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc699     Doc700     Doc701     Doc702     Doc703     Doc704     Doc705 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc706     Doc707     Doc708     Doc709     Doc710     Doc711     Doc712 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc713     Doc714     Doc715     Doc716     Doc717     Doc718     Doc719 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc720     Doc721     Doc722     Doc723     Doc724     Doc725     Doc726 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc727     Doc728     Doc729     Doc730     Doc731     Doc732     Doc733 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc734     Doc735     Doc736     Doc737     Doc738     Doc739     Doc740
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc741     Doc742     Doc743     Doc744     Doc745     Doc746     Doc747
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc748     Doc749     Doc750     Doc751     Doc752     Doc753     Doc754
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc755     Doc756     Doc757     Doc758     Doc759     Doc760     Doc761
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc762     Doc763     Doc764     Doc765     Doc766     Doc767     Doc768
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc769     Doc770     Doc771     Doc772     Doc773     Doc774     Doc775
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc776     Doc777     Doc778     Doc779     Doc780     Doc781     Doc782
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc783     Doc784     Doc785     Doc786     Doc787     Doc788     Doc789
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc790     Doc792     Doc793     Doc794     Doc795     Doc796     Doc797
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc798     Doc799     Doc800     Doc801     Doc802     Doc803     Doc804
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc805     Doc806     Doc807     Doc808     Doc809     Doc810     Doc811
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc812     Doc813     Doc814     Doc815     Doc816     Doc817     Doc818
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc819     Doc820     Doc821     Doc822     Doc823     Doc824     Doc825
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc826     Doc827     Doc828     Doc829     Doc830     Doc831     Doc832
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc833     Doc834     Doc835     Doc836     Doc837     Doc838     Doc839
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc840     Doc841     Doc842     Doc843     Doc844     Doc845     Doc846
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc847     Doc848     Doc849     Doc850     Doc851     Doc852     Doc853
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc854     Doc855     Doc856     Doc857     Doc858     Doc859     Doc860
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc861     Doc862     Doc863     Doc864     Doc865     Doc866     Doc867
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc868     Doc869     Doc870     Doc871     Doc872     Doc873     Doc874
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc875     Doc876     Doc877     Doc878     Doc879     Doc880     Doc881
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc882     Doc883     Doc884     Doc885     Doc886     Doc887     Doc888
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc889     Doc890     Doc891     Doc892     Doc893     Doc894     Doc895
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc896     Doc897     Doc898     Doc899     Doc900     Doc901     Doc902
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc903     Doc904     Doc905     Doc906     Doc907     Doc908     Doc909
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc910     Doc911     Doc912     Doc913     Doc914     Doc915     Doc916
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc917     Doc918     Doc919     Doc920     Doc921     Doc922     Doc923
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##     Doc924     Doc925     Doc926     Doc927     Doc928     Doc929     Doc930 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc931     Doc932     Doc933     Doc934     Doc935     Doc936     Doc937 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc938     Doc939     Doc940     Doc941     Doc942     Doc943     Doc944 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc945     Doc946     Doc947     Doc948     Doc949     Doc950     Doc951 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc952     Doc953     Doc954     Doc955     Doc956     Doc957     Doc958 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc959     Doc960     Doc961     Doc962     Doc963     Doc964     Doc965 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc966     Doc967     Doc968     Doc969     Doc970     Doc971     Doc972 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc973     Doc974     Doc975     Doc976     Doc977     Doc978     Doc979 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc980     Doc981     Doc982     Doc983     Doc984     Doc985     Doc986 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc987     Doc988     Doc989     Doc990     Doc991     Doc992     Doc993 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 
##     Doc994     Doc995     Doc996     Doc997     Doc998     Doc999 
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
```

```
#heatmap not informative, too many datapoints

#heatmap(c.similarity.matrix.uni[, ])
#heatmap(c.similarity.matrix.bi[, ])

#pairheatmap(c.similarity.matrix.uni[,], c.similarity.matrix.bi[,], colorStyle="s3")
```

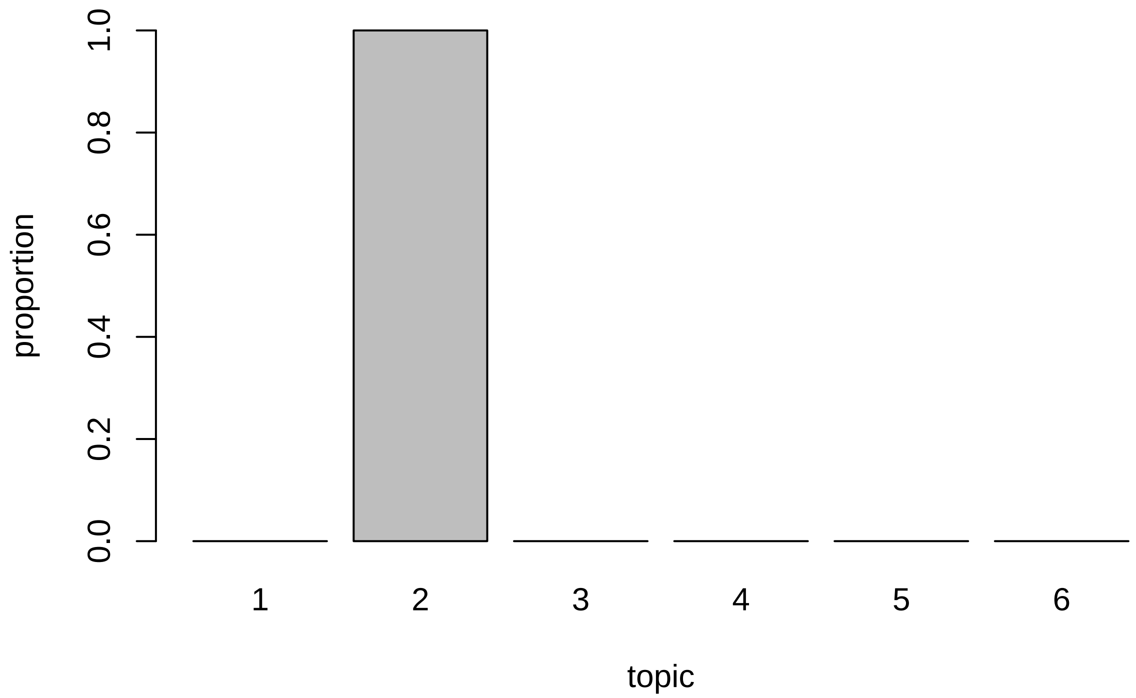
```
#LDA - topic relevance
dtm = as(term.doc.matrix.bigram, "dgTMatrix")

lda_model = text2vec::LDA$new(n_topics = 6, doc_topic_prior = 0.1, topic_word_prior = 0.01)

doc_topic_distr = lda_model$fit_transform(x = dtm, n_iter = 1000,
                          convergence_tol = 0.001, n_check_convergence = 25,
                          progressbar = TRUE)
```

```
##   |                                                                               |
```

```
#plotting results
barplot(doc_topic_distr[1, ], xlab = "topic",
        ylab = "proportion", ylim = c(0, 1),
        names.arg = 1:ncol(doc_topic_distr))
```



```
lda_model$get_top_words(n = 6, topic_number = c(1L, 3L, 6L), lambda = 1)
```

```
##      [,1]  [,2]  [,3]
## [1,] "93"  "963" "277"
## [2,] "861" "85"  "711"
## [3,] "936" "235" "334"
## [4,] "263" "606" "564"
## [5,] "742" "409" "947"
## [6,] "113" "637" "53"
```

```
lda_model$get_top_words(n = 6, topic_number = c(1L, 3L, 6L), lambda = 0.2)
```

```
##      [,1]  [,2]  [,3]
## [1,] "93"  "963" "277"
## [2,] "861" "85"  "711"
## [3,] "263" "235" "564"
## [4,] "742" "606" "947"
## [5,] "113" "409" "53"
## [6,] "220" "637" "374"
```

```
lda_model$plot()
```

```
## Loading required namespace: servr
```